WHAT IS CLAIMED IS:

1. A method for compressing an index file in an information retrieval system that retrieves information from a plurality of documents, each of the plurality of documents having features occurring therein, the method comprising the step of:

representing occurrence frequencies of the features in the plurality of documents in a compressed format in the index file.

2. The method of claim 1, wherein the features are textual.

3. The method of claim 1, wherein the features are non-textual.

4. The method of claim 1, wherein said representing step comprises the steps of:

mapping the occurrence frequencies into a plurality of bins; and

storing bin identifiers in the index file, each of the bin identifiers identifying a bin to which at least one individual occurrence frequency is assigned.

5. The method of claim 4, further comprising the step of establishing each of the plurality of bins to represent a numerical interval that contains at least one of the occurrence frequencies.

6. The method of claim 5, wherein at least one of the plurality of bins represents an empty bin.

7. The method of claim 4, further comprising the step of establishing each of the plurality of bins to represent a different numerical interval, such that the different numerical interval represented by each of the plurality of bins contains a substantially same number of the occurrence frequencies.

8. The method of claim 4, wherein said mapping step respectively maps more than a single term and a corresponding occurrence frequency into each of the plurality of bins, the method further comprises the step of scoring at least one of the plurality of documents with respect to a query, and said scoring step comprises the step of computing an occurrence frequency for a given one of the

plurality of bins as a weighted average of the occurrence frequencies contained within the given one of the plurality of bins.

9. The method of claim 4, wherein said mapping step respectively maps only a single term and a corresponding occurrence frequency into each of the plurality of bins, the method further comprises the step of scoring at least one of the plurality of documents with respect to a query, and said scoring step comprises the step of computing an occurrence frequency for a given one of the plurality of bins based on the single term and corresponding occurrence frequency mapped thereto.

10. The method of claim 4, further comprising the step of establishing bin boundaries for the plurality of bins based on a methodology employed to score the plurality of documents with respect to queries, the bin boundaries defining intervals within which the occurrence frequencies fall.

11. The method of claim 4, further comprising the steps of:

receiving a query having at least one term; and

computing a relevance score for at least one of the plurality of documents with respect to the query, based on the bin identifiers.

12. The method of claim 4, wherein the method is implemented by a program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform said method steps.

13. An apparatus for compressing an index file in an information retrieval system that retrieves information from a plurality of documents, each of the plurality of documents having features occurring therein, the apparatus comprising:

a compression device for representing occurrence frequencies of the features in the plurality of documents in a compressed format in the index file.

14. The apparatus of claim 13, wherein the features are textual.

15. The apparatus of claim 13, wherein the features are non-textual.

16. The apparatus of claim 13, wherein said compression device comprises:

a bin generator for generating a plurality of bins and a corresponding plurality of bin identifiers, each of the plurality of bin identifiers respectively identifying one of the plurality of bins to which at least one individual occurrence frequency is mapped;

a mapping device for mapping the occurrence frequencies into the plurality of bins; and

a storage device for storing the bin identifiers in the index file.

17. The apparatus of claim 16, wherein said bin generator establishes each of the plurality of bins to represent a numerical interval that contains at least one of the occurrence frequencies.

18. The apparatus of claim 16, wherein said bin generator establishes each of the plurality of bins to represent a different numerical interval, such that the

different numerical interval represented by each of the plurality of bins contains a substantially same number of the occurrence frequencies.

19. The apparatus of claim 16, wherein said mapping device respectively maps more than a single term and a corresponding occurrence frequency into each of the plurality of bins, the apparatus further comprises a scoring device for scoring at least one of the plurality of documents with respect to a query by computing an occurrence frequency for a given one of the plurality of bins as a weighted average of the occurrence frequencies contained within the given one of the plurality of bins.

20. The apparatus of claim 16, wherein said mapping device respectively maps only a single term and a corresponding occurrence frequency into each of the plurality of bins, the apparatus further comprises a scoring device for scoring at least one of the plurality of documents with respect to a query by computing an occurrence frequency for a given one of the plurality of bins based on the single term and corresponding occurrence frequency mapped thereto.

21. The apparatus of claim 16, wherein said bin generator establishes bin boundaries for the plurality of bins based on a methodology employed to score the plurality of documents with respect to queries, the bin boundaries defining intervals within which the occurrence frequencies fall.

22. The apparatus of claim 16, further comprising a scoring device for computing a relevance score for at least one of the plurality of documents with respect to a query, based on the bin identifiers.

23. A method for compressing an index file in an information retrieval system that retrieves information from a plurality of documents, each of the plurality of documents having features occurring therein, each of the features having parameters corresponding thereto, the method comprising the step of:

mapping parameter values corresponding to the parameters of the features into a plurality of bins; and

storing bin identifiers in the index file, each of the bin identifiers identifying a bin to which is assigned at

least one individual parameter value corresponding to at least one individual parameter.

24. The method of claim 23, wherein the features are textual.

25. The method of claim 23, wherein the features are non-textual.

26. The method of claim 23, further comprising the step of establishing each of the plurality of bins to represent a numerical interval that contains at least one of the parameter values.

27. The method of claim 26, wherein at least one of the plurality of bins represents an empty bin.

28. The method of claim 23, further comprising the step of establishing each of the plurality of bins to represent a different numerical interval, such that the different numerical interval represented by each of the plurality of bins contains a substantially same number of the parameter values.

29. The method of claim 23, wherein said mapping step respectively maps more than a single parameter and a corresponding parameter value into each of the plurality of bins, the method further comprises the step of scoring at least one of the plurality of documents with respect to a query, and said scoring step comprises the step of computing a parameter value for a given one of the plurality of bins as a weighted average of the parameter values contained within the given one of the plurality of bins.

30. The method of claim 23, wherein said mapping step respectively maps only a single parameter and a corresponding parameter value into each of the plurality of bins, the method further comprises the step of scoring at least one of the plurality of documents with respect to a query, and said scoring step comprises the step of computing a parameter value for a given one of the plurality of bins based on the single parameter and corresponding parameter value mapped thereto.

31. The method of claim 23, further comprising the step of establishing bin boundaries for the plurality of bins based on a methodology employed to score the plurality

of documents with respect to queries, the bin boundaries defining intervals within which the parameter values fall.

32. The method of claim 23, further comprising the steps of:

receiving a query having at least one parameter; and

computing a relevance score for at least one of the plurality of documents with respect to the query, based on the bin identifiers.

33. The method of claim 23, wherein the method is implemented by a program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform said method steps.